Mrs. Swati Amit Motugade

MACHINE LEARNING WORKSHEET 6

Q.1. B

Q.2. B

Q.3. C

Q.4. A

Q.5. B

Q.6. A & D

Q.7. B & C

Q.8. A & C

Q.9. A & B

Q.10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

ANS. The adjusted R^2 tells you the percentage of variation explained by only the independent variables that actually affect the dependent variable. The adjusted R^2 will penalize you for adding independent variables (K in the equation) that do not fit the model. Because, in regression analysis, it can be tempting to add more variables to the data as you think of them. Some of those variables will be significant, but you can't be sure that significance is just by chance.

Q. 11. Differentiate between Ridge and Lasso Regression.

Ans.

1. Ridge Regression:

In Ridge regression, you estimate the coefficients by trying to minimize the sum of squares of errors (this focuses on prediction accuracy, just like normal multiple linear regression) with the constraint that the sum of the squares of the coefficients is less than a certain value (this constraint forms the L2 Norm and keeps the coefficients from growing large). The constraint, in other words, penalizes the coefficients. When you dive deeper into Ridge

regression, you will see that L2 regularization doesn't necessarily push the coefficients to zero. What this means is that Ridge regression is useful in reducing or preventing overfitting, but may not really help with model parsimony or with feature selection.

2. Lasso Regression:

In Lasso, what changes is that the constraint is now to keep the sum of the absolute values of the coefficients less than a certain value. Just like in Ridge regression, the constraint penalizes the coefficients and keeps them from growing large, thus reducing or preventing overfitting. However, when you dive into the mathematics, you'll see that Lasso, compared to Ridge, can actually push the coefficients to zero, thereby removing variables / features from the model, thus achieving model parsimony and feature selection as well.

Q. 12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans. A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis. Multicollinearity exists when there is a correlation between multiple independent variables in a multiple regression model. This can adversely affect the regression results. Thus, the variance inflation factor can estimate how much the variance of a regression coefficient is inflated due to multicollinearity.

A variance inflation factor (VIF) provides a measure of multicollinearity among the independent variables in a multiple regression model.

A large VIF on an independent variable indicates a highly collinear relationship to the other variables that should be considered or adjusted for in the structure of the model and selection of independent variables.

As a rule of thumb, a VIF of three or below is not a cause for concern. As VIF increases, the less reliable your regression results are going to be.

Q. 13. Why do we need to scale the data before feeding it to the train the model?

Ans. Feature scaling is essential for machine learning algorithms that calculate distances between data. If not scale, the feature with a higher value range starts dominating when calculating distances.

Scaling of the data comes under the set of steps of data pre-processing when we are performing machine learning algorithms in the data set. As we know most of the supervised and unsupervised learning methods make decisions according to the data sets applied to them and often the algorithms calculate the distance between the data points to make better inferences out of the data.

In real life, if we take an example of purchasing apples from a bunch of apples, we go close to the shop, examine various apples and pick various apples of the same attributes. Because we have learned about the attributes of apples and we know which are better and which are not good also we know which attributes can be compromised and which cannot. So, if most of the apples consist of pretty similar attributes we will take less time in the selection of the apples which directly affect the time of purchasing taken by us. The moral of the example is if the apples every apple in the shop is good, we will take less time to purchase or if the apples are not good enough, we will take more time in the selection process which means that if the values of attributes are closer, we will work faster and the chances of selecting good apples also strong.

Similarly in the machine learning algorithms if the values of the features are closer to each other there are chances for the algorithm to get trained well and faster instead of the data set where the data points or features values have high differences with each other will take more time to understand the data and the accuracy will be lower.So if the data in any conditions has data points far from each other, scaling is a technique to make them closer to each other or in simpler words, we can say that the scaling is used for making data points generalized so that the distance between them will be lower.

Q. 14. What are the different metrics which are used to check the goodness of fit in linear regression?

Ans. The Mean Squared Error, Root Mean Squared Error and Mean Absolute Error are the metrics used to check the goodness of fit in linear regression.

1. Mean Squared Error:

The Mean Squared Error or MSE is a popular error metric for regression problems. It is also an important loss function for algorithms fit or optimized using the least squares framing of a regression problem. Here “least squares”

refers to minimizing the mean squared error between predictions and expected values.

The MSE is calculated as the mean or average of the squared differences between predicted and expected target values in a dataset. The units of the MSE are squared units.

2. Root Mean Squared Error:

Root Mean Squared Error or RMSE is an extension of the mean squared error.

The square root of the error is calculated, which means that the units of the RMSE are the same as the original units of the target value that is being predicted.

For example, if your target variable has the units “dollars,” then the RMSE error score will also have the unit “dollars” and not “squared dollars” like the MSE.

As such, it may be common to use MSE loss to train a regression predictive model, and to use RMSE to evaluate and report its performance.

The RMSE can be calculated as follows:

RMSE = sqrt (MSE)

3. Mean Absolute Error:

Mean Absolute Error or MAE is a popular metric because, like RMSE, the units of the error score match the units of the target value that is being predicted.

Unlike the RMSE, the changes in MAE are linear and therefore intuitive.

That is, MSE and RMSE punish larger errors more than smaller errors, inflating or magnifying the mean error score. This is due to the square of the error value. The MAE does not give more or less weight to different types of errors and instead the scores increase linearly with increases in error.

As its name suggests, the MAE score is calculated as the average of the absolute error values. Absolute or abs () is a mathematical function that simply makes a number positive. Therefore, the difference between an expected and

predicted value may be positive or negative and is forced to be positive when calculating the MAE.

Q. 15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Actual/Predicted	True	False
True	1000	50
False	250	1200

Ans.

1. Precision = TP / (TP + FP) = 1000/ (1000+50) = 0.952380

2. Sensitivity/ Recall = TP / (TP + FN) = 1000 / (1000+250) = 0.8

3. Specificity = TN / (TN + FP) = 1200/ (1200+50) = 0.96

4. Accuracy = (TP + TN) / (TP+TN+PF+FN) = (1000+1200)/(1000+1200+50+250) = 0.88